

## Research article

## Evolution of secretin family GPCR members in the metazoa

João CR Cardoso<sup>1</sup>, Vanda C Pinto<sup>1</sup>, Florbela A Vieira<sup>1</sup>, Melody S Clark<sup>2</sup> and Deborah M Power<sup>\*1</sup>

Address: <sup>1</sup>Centre of Marine Sciences, Universidade do Algarve, Campus de Gambelas, 8005-139 Faro, Portugal and <sup>2</sup>British Antarctic Survey, High Cross, Madingley Road, Cambridge, CB3 0ET, Cambridge, UK

Email: João CR Cardoso - [jccardo@ualg.pt](mailto:jccardo@ualg.pt); Vanda C Pinto - [vanda\\_pinto@hotmail.com](mailto:vanda_pinto@hotmail.com); Florbela A Vieira - [fvieira@ualg.pt](mailto:fvieira@ualg.pt); Melody S Clark - [mscl@bas.ac.uk](mailto:mscl@bas.ac.uk); Deborah M Power\* - [dpower@ualg.pt](mailto:dpower@ualg.pt)

\* Corresponding author

Published: 13 December 2006

Received: 09 August 2006

BMC Evolutionary Biology 2006, 6:108 doi:10.1186/1471-2148-6-108

Accepted: 13 December 2006

This article is available from: <http://www.biomedcentral.com/1471-2148/6/108>

© 2006 Cardoso et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

**Background:** Comparative approaches using protostome and deuterostome data have greatly contributed to understanding gene function and organismal complexity. The family 2 G-protein coupled receptors (GPCRs) are one of the largest and best studied hormone and neuropeptide receptor families. They are suggested to have arisen from a single ancestral gene via duplication events. Despite the recent identification of receptor members in protostome and early deuterostome genomes, relatively little is known about their function or origin during metazoan divergence. In this study a comprehensive description of family 2 GPCR evolution is given based on *in silico* and expression analyses of the invertebrate receptor genes.

**Results:** Family 2 GPCR members were identified in the invertebrate genomes of the nematodes *C. elegans* and *C. briggsae*, the arthropods *D. melanogaster* and *A. gambiae* (mosquito) and in the tunicate *C. intestinalis*. This suggests that they are of ancient origin and have evolved through gene/genome duplication events. Sequence comparisons and phylogenetic analyses have demonstrated that the immediate gene environment, with regard to gene content, is conserved between the protostome and deuterostome receptor genomic regions. Also that the protostome genes are more like the deuterostome Corticotrophin Releasing Factor (CRF) and Calcitonin/Calcitonin Gene-Related Peptide (CAL/CGRP) receptors members than the other family 2 GPCR members. The evolution of family 2 GPCRs in deuterostomes is characterised by acquisition of new family members, with SCT (Secretin) receptors only present in tetrapods. Gene structure is characterised by an increase in intron number with organismal complexity with the exception of the vertebrate CAL/CGRP receptors.

**Conclusion:** The family 2 GPCR members provide a good example of gene duplication events occurring in tandem with increasing organismal complexity during metazoan evolution. The putative ancestral receptors are proposed to be more like the deuterostome CAL/CGRP and CRF receptors and this may be associated with their fundamental role in calcium regulation and the stress response, both of which are essential for survival.

## Background

The Guanine protein coupled receptor (GPCRs) family is one of the largest receptor groups in vertebrates. Members of this family are also present in unicellular eukaryotes such as yeast and in plants which suggests that they are of ancient origin [1]. In the human genome, GPCRs account for approximately 2% of the coding genes [2,3] and they bind structurally diverse ligands such as protons, odorants, biogenic amines, peptides and glycoproteins [4,5]. Recent analysis of the human genome identified five main GPCR subfamilies collectively known as GRAFS (Glutamate, G; Rhodopsin, R; Adhesion, A; Frizzled, F; and Secretin, S) [1,6]. This grouping was based on protein motifs characterised by the presence of seven highly conserved transmembrane domains (TM). Several authors have proposed the existence of a common ancestral gene in early metazoans that, as a consequence of successive duplication events, generated the full complement of family members in vertebrates [6,7]. Such a proposal is in general agreement with the genome duplication theories of Haldane 1932 [8], Muller 1935 [9] and Ohno 1970 [10]. All of which suggest that the existence of gene family members in chordates is a consequence of genome duplication events in the vertebrate lineage during evolution and that gene duplicates are an essential source of organism diversity.

The present work focuses on the secretin family (a.k.a family B or 2) of GPCRs which represent one of the largest receptor families for hormones and neuropeptides involved in several important biological functions. Previous *in silico* analysis identified a total of 50 family 2 GPCR members in the human genome [6,11,12]. However only receptor members of the following groups: a) Corticotrophin Releasing Factor (CRF); b) Secretin (SCT), Vasoactive Intestinal Peptide (VIP), Pituitary Adenylate Cyclase-Activating Polypeptide (PACAP) and Growth Hormone Releasing Hormone (GHRH); c) Glucagon (GCG), Glucagon-Like Peptide (GLP), Glucose Insulinotropic Peptide (GIP); d) Parathyroid Hormone (PTR) and e) Calcitonin (CAL) and Calcitonin Gene-Related Peptide (CGRP) [13,14] have been functionally characterised and identified in other vertebrates such as birds, amphibians and teleosts [15-18]. So far, in invertebrates putative family 2 GPCRs have been found based upon sequence similarity and phylogenetic studies ([1,14,17,19,20], but their evolution and function is poorly described.

In this study, comparative analyses using phylogenetically distant organisms have been used to study the evolution of members of the family 2 GPCRs in metazoans (Figure 1). Gene family members were characterised in protostome (nematode and arthropod) and tunicate (*Ciona*) genomes and compared with their vertebrate homologues (*Takifugu* and human). Putative family 2 GPCR receptor

members were identified and isolated *in silico* from public genome databases and their expression analysed by RT-PCR. The gene structures and gene environments *vis à vis* gene content of the protostome and deuterostome receptors were compared and a model for the evolution of family 2 GPCR receptors is proposed.

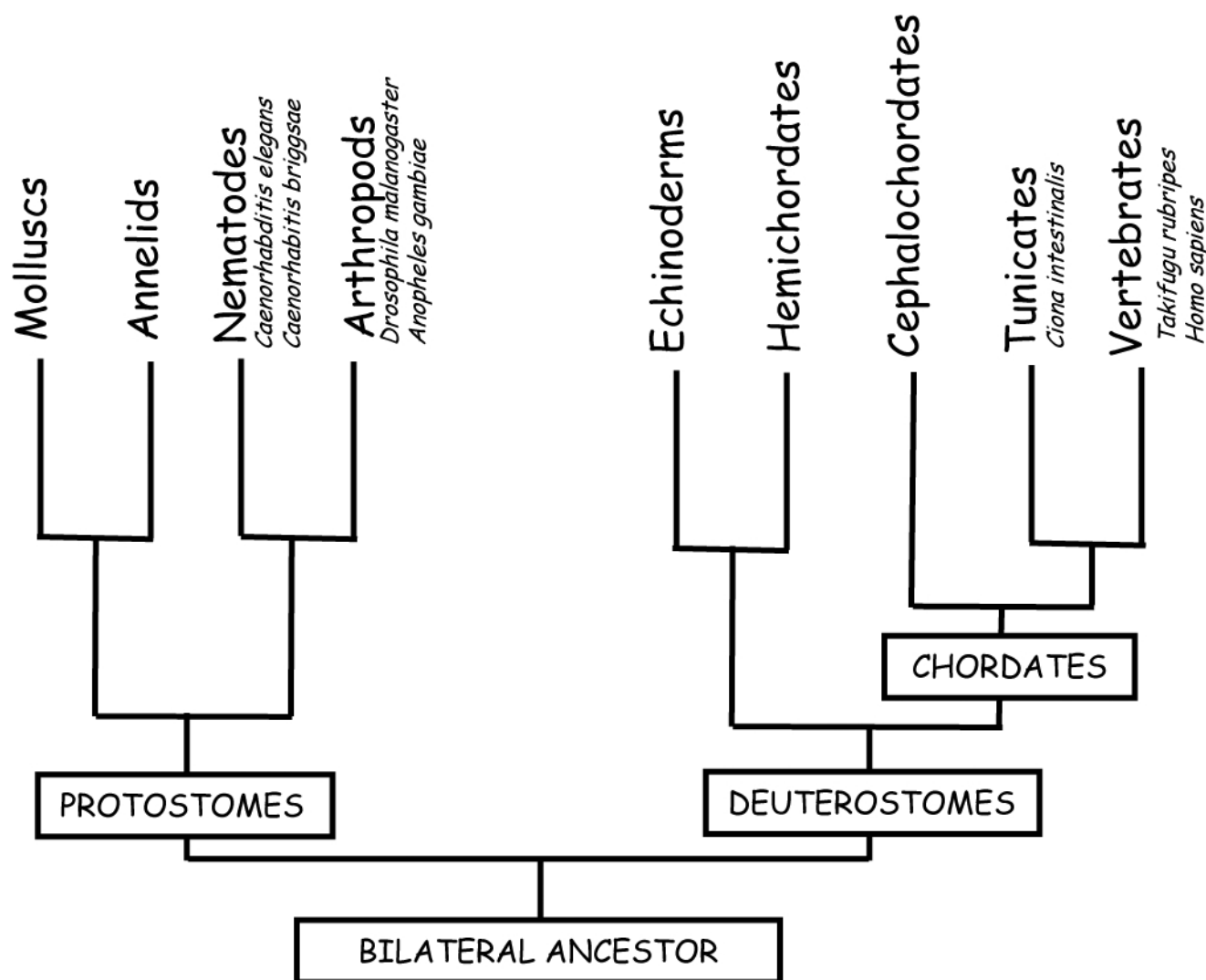
## Results

### Putative family 2 GPCRs in invertebrates

Initially, three, five and nine putative family 2 GPCR genes were identified respectively in the protostome genomes of the nematodes (*C. elegans* and *C. briggsae*), the arthropods (*D. melanogaster* and *Anopheles gambiae*) and in the deuterostome genome of the tunicate *C. intestinalis*. With the exception of the *C. briggsae* and mosquito receptor genes, the protostome and tunicate predicted gene sequences were edited taking into consideration EST data available in order to minimise errors derived from *in silico* gene predictions (Additional file 1). Sequence comparison and other *in silico* approaches revealed that of these genes, only two in the nematodes, three in the fruit-fly, one gene in the mosquito and eight in the *Ciona* genome contained seven TM domains and were considered to be putative family 2 GPCRs members (Figure 2, Table 1). Database searches carried out in the prokaryote *E. coli* and the unicellular eukaryote *S. cerevisiae* genomes did not produce any significant alignments for family 2 receptor genes.

### Phylogenetic analysis

Because of gaps and errors in draft genome data, only 4 TMs were used in the full analysis: TM2, TM4, TM5 and TM6 which were found to be common to all metazoan genes analysed (Additional file 2). The amino acid sequences of the TM domains of a total of 52 receptors were concatenated and aligned using the ClustalX programme (Blosom matrix and Gap opening penalty 10 and Gap extension 0.2). The alignment produced (length 98, with 99 informative sites) did not required the insertion of gaps and was used for phylogenetic analysis and the consensus tree obtained is presented on Figure 2. Family 2 GPCR members are suggested to have evolved via both late and early gene duplication events, which have occurred during metazoan evolution. Examples of specific gene duplication events which are well supported by the high bootstrap values of the tree nodes are the *Ciona* gene pair CinS5A/CinS5B, CinS372/CinS752 and the gene pair CinS50/CinS273. After neighbour joining, maximum parsimony and minimum evolution phylogenetic analysis, the protostome receptors and *Ciona* CinS93, CinS50 and CinS273 genes tended to cluster with the deuterostome CRF and CAL/CGRP receptor family members. In *Ciona* orthologues of the majority of the vertebrate family members such as CinS752 and CinS352 group with the PTR receptor subfamily, whilst CinS5A, CinS5B and CinS70

**Figure 1**

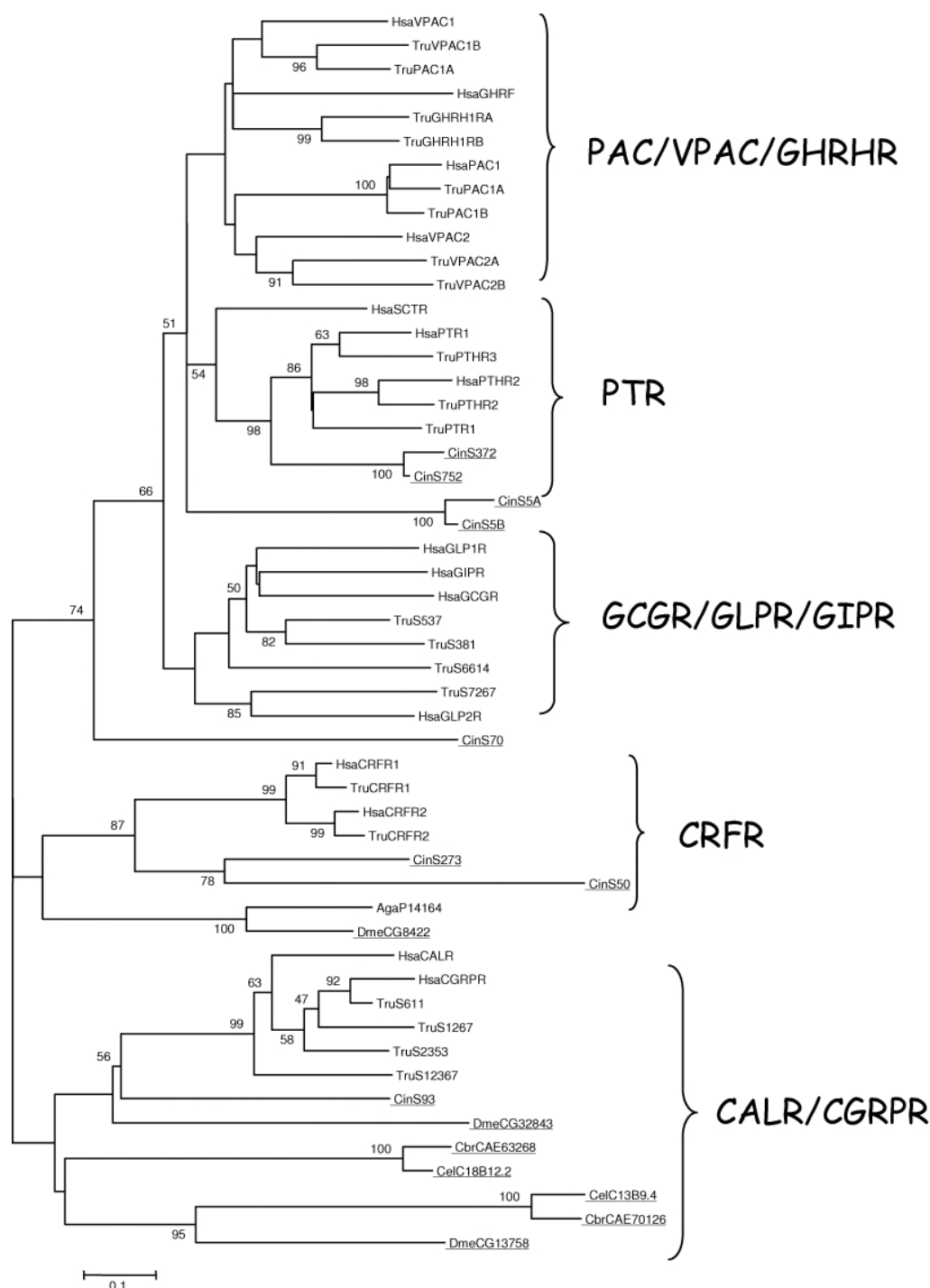
**Phylogenetic position of the protostome and deuterostome genomes analysed.** Simplified phylogeny of the metazoan evolution based on molecular data indicating the positions of protostome (nematodes and arthropods) and deuterostome (tunicate and vertebrate) genomes analysed (adapted from Giribet., 2002 [79]; Gerhart et al., 2005 [80]; Delsuc et al., 2006 [81]).

genes seem to be more related in sequence to the GCG/GLP/GIP and PAC/VPAC/GHRH families however CinS5A and CinS5B position in the tree was not clearly defined.

#### Sequence comparative analyses

Comparison of the amino acid sequences of the putative protostome and tunicate receptors with the vertebrate homologues revealed the existence of conserved amino acid motifs at the ligand-binding N-terminal region (Figure 3). Large N-terminal regions containing five conserved

cysteines were identified in all species analysed with the exception of the *Ciona* CinS273 and CinS70 genes which due to incomplete genome coverage and EST data available lacked complete N-terminal regions and CinS50 where the putative initial methionine was not identified. Other highly conserved amino acid residues such as the amino acid aspartate (D) before the motif C-W-P and the amino acid motifs C-W-P, C-P and G-X-W (where X is any amino acid) (crucial for ligand binding in mammals) [21,22] were also identified. Moreover, the amino acid glycine (G) localised between this latter motif and the C-

**Figure 2**

**Phylogenetic relationship of the metazoan family 2 GPCRs.** Consensus phylogenetic tree (neighbour joining method, pairwise gap deletion, Poisson correction distance and 1000 bootstraps) produced with family 2 GPCR TM domains (TM2, TM4, TM5 and TM6). The protostome (nematodes and arthropods) and tunicate (*Ciona*) receptors are underlined and the bootstrap values for each fork is indicated. Bootstrap values less than 50 were removed. Annotation of the receptor sub-families was carried out according to Donnelly (1997) [13]. The human (Hsa), *Takifugu* (Tru), nematodes, *C. elegans* (Cel) and *C. briggsae* (Cbr), arthropod *D. melanogaster* (Dme) and *A. gambiae* (Aga) and *Ciona* (Cin) receptor sequences were obtained as described in the methods section.

**Table 1: List of putative invertebrate family 2 GPCRs identified**

Nematodes		Arthropods		Tunicate
<i>C. elegans</i> (Cel)	<i>C. briggsae</i> (Cbr)	<i>D. melanogaster</i> (Dme)	<i>A. gambiae</i> (Aga)	<i>C. intestinalis</i> (Cin)
CelC18B12.2	CbrCAE63268	DmeCG13758	AgaENSANGP00000014363*	CinS93
CelC13B9.4	CbrCAE70126	DmeCG8422	AgaENSANGP00000014114+	CinS2303+
CelZK643.3 *	CbrCAE62707*	DmeCG12370*	AgaENSANGP00000020176*	CinS70
		DmeCG32843	AgaENSANGP00000014164	CinS5 (A and B)
		DmeCG4395*#	AgaENSANGP00000004125#	CinS273
				CinS752
				CinS372
				CinS50

Names of the protostomes and tunicate receptor clones identified by sequence similarity searches using the vertebrate family 2 GPCRs (\*) does not have 7 TM domains; (#) TM4 was not identified; (+) incomplete

P motif was also found to be conserved amongst metazoans however no functional role for this residue has been yet assigned.

To strengthen previous analysis and in order to identify novel amino acid and protein motifs that have been conserved within each receptor subfamily, members that might have a potential role in ligand binding in the protostome and deuterostome N-terminal regions of the receptors were compared based on receptor clustering groups previously obtained by phylogenetic analysis. Putative protostome and tunicate members of the CAL/CGRP, CRF, PTH and CGC/GLP/GIP receptor families were aligned with the vertebrate homologue receptor genes (Additional files 3456). Examples of novel conserved N-terminal protein motifs identified are the W-S/T-N-Y/F motif in CAL/CGRP receptors alignment (Additional file 3); the motif G-V/I-X-Y (X any amino acid) within CRF receptor group (Additional file 4) which has been previously reported to be involved in ligand binding [23,24]; the motifs E-W, P-G; Y-I-Y/I-D-F-D/N-H and A-X-R (X any amino acid) in the PTR group and amino acid HsaPTR1 R<sup>186</sup>(Additional file 5) which was previously found to be determinant for PTH binding [25] and the motif Y-L/I-P/E-W within GCG/GLP/GIP receptor group (Additional file 6).

#### Short-range linkage mapping analysis

Short-range linkage analysis was carried out between the protostome (*C. elegans* and *D. melanogaster*) with the deuterostome (*Takifugu* and human) homologous regions which contain family 2 receptor genes. The linked genes identified in both *C. elegans* and *Takifugu* were used to identify homologous genes in the *Drosophila* (Figure 4) and human (Figure 5) genomes. The protostome gene environment was found to be conserved and linked genes were identified between the nematode and insect genomes. The *Drosophila* X chromosome and chromosome III of *C. elegans* showed the greatest number of linked genes. Within the protostome genome three genes

namely 3H538, Clp-2 and him-4 in *C. elegans* were also found to be conserved in the homologous deuterostomes genome regions analysed (Figure 5). A number of *Takifugu* scaffolds were found to share a similar gene environment with *C. elegans* chromosomes II and X, and human chromosomes 2, 3, 6, 7, and 17. This suggests that the nematode chromosome regions may be very similar in terms of gene content with the ancestral chromosomal region that gave rise to this family of receptors in vertebrates (Figure 5).

#### Gene organisation in protostomes and deuterostomes

The gene organisation of the regions encompassing the seven TM domains of the protostome and tunicate family 2 GPCR genes were characterised and compared with the vertebrate human and *Takifugu* homologous regions (Figure 6). Different gene structures were observed between protostomes and deuterostomes. This was mainly due to an increase in intron number in the latter species. Comparison of exon/intron boundaries revealed total conservation of splice sites (AG/GT) although intron phases are generally poorly conserved with the exception of the TM1 and TM2 boundary (Figure 6).

Amongst protostome family 2 GPCR genes, organisation is poorly conserved. In nematodes two different gene organisations were identified for the receptors CelC13B9.4/CbrCAE70126 (5 exons) and CelC18B12.2/CbrCAE63268 (6 exons), respectively. In the *Drosophila* genome, a different gene organisation also exists for each of the three receptor genes analysed. TM1 and TM2 in the *Drosophila* genes DmeCG32843 and DmeCG8422, in common with the nematode, tunicate and vertebrate receptors, are encoded by individual exons. The DmeCG13758 gene has the most divergent gene structure amongst protostomes and is composed of 3 exons. In mosquito, the gene structure of AgENSANGP00000014164 (AgaP14164) is similar to DmeCG8422 with which it share greatest sequence similarity.

HsaCALR	:	-----MRFTFTSRCLALFLLLNHPTILPAFSNQTYPTIEPKFLYVVRKKMM	:	49
HsaGLP1R	:	-----MAGAPGLRLALLLLGMVGRAGPRPQGATVSLWETVQKWR	:	40
HsaCRFR1	:	-----MGGHPQLRLVKALLLLGLNPVSAS	:	24
CinS5A	:	-----MCKYM	:	5
CinS5B	:	-----MCKYM	:	5
CinS93	:	-----MYNFEEFT	:	7
CinS372	:	-----MADDGITTLOQVMEVLL	:	17
CinS752	:	-----MADDGITTREMMEVLM	:	17
AdoDHR	:	-----MEDAWEADALEAAEAAAAAQAADAPTTPSPRTL	:	37
DmeCG8422	:	-----MSDHNHIDSVNASGSDPLLDLHNLGDIGE	:	29
DmeCG13758	:	MTLLSNILDCGGCISAQRFTRLLRQSGSSGSPSAPTAGTFESKSMLEPTSSHSLATGRVPLLDHFDASTTESPGTYVLD	:	80
DmeCG32843	:	-----MSDQIGNPNATFSGSGSGSGTNVASIAESVAESGPDFD	:	38
AgaP14164	:	-----MANETMPTTPTAAIAADDGSGSGSPFGEENVTGISVENPSLV	:	42
CelC13B9	:	-----MADATSPFNVSILDN-STKLSEMVESGWNVLASTSVQAFNEAMDVLE	:	46
CelC18B12	:	-----MTSIIIR	:	6
CbrCAE6326	:	-----MTSIIQ	:	6
CbrCAE7012	:	-----MADATSPFNVSILDNSTAKLSEMVDSGWNVLASTSVQAFNGAMDVLE	:	47

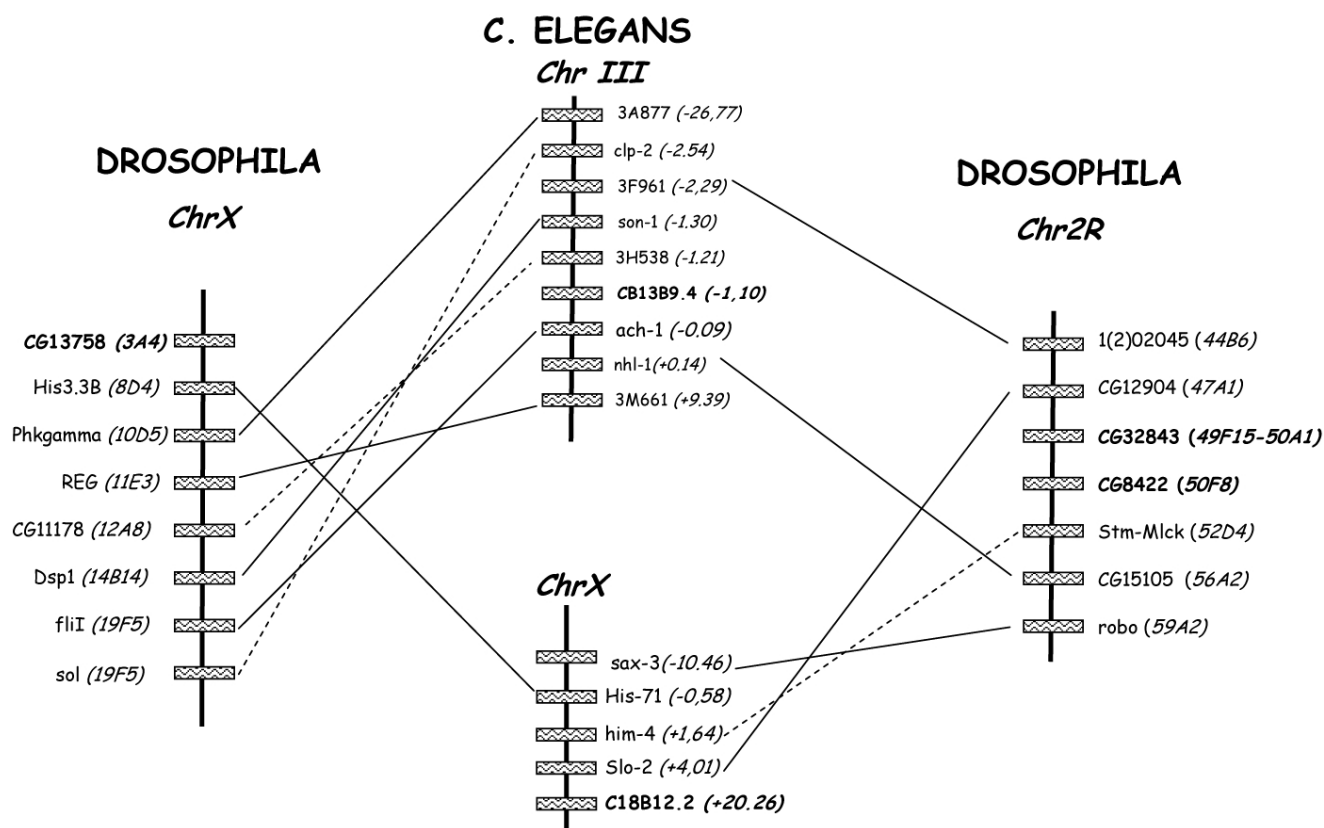
HsaCALR	:	DAQYKCYDRMQQLP-----AYQGEQGYCNRTWDEW-LCWDPTAGVLSYQFCPDYFPDFDPSEKVT-----KYCD	:	113
HsaGLP1R	:	EYRRQCQRSLTED-----PPPATDLFCNRTFDEY-ACWPDGEPGGFVNVSCPWYLPWASSVPQGH-----VYRFT	:	105
HsaCRFR1	:	LQDQHCELSLAS-----NISGLQCNASVDLIGTCWPRSPAGQLVVRPCPAFFYGVRYNTTNN-----GYRECL	:	88
CinS5A	:	QYRERCEKSLNETLKN-----FENQTFPCGIFDGY-ICWPHTPAGQLLTERCPGYIEWKN-FESLV-----AYRCE	:	70
CinS5B	:	KYKEQCEKSLNETLKN-----FENQTFPCGIFDQY-ICWPHTPAGQLLTERCPGYIEWR-NESLV-----AIRCE	:	69
CinS93	:	RQFNCFCVNDNLNISRNSK-----SNEFVCSFCGAIWDGY-ACWPPQTSSKEAVQNCPSYLPFRNTEEFVAVR-----PCQ	:	75
CinS372	:	EAQRNCKRDLEA-----GLPPKDKDGEWDKI-TCPWNSAPGRTVRLFCPEYIIDF--DHTGH-----ALRHCS	:	77
CinS752	:	EAQRNCKRDLEA-----GLPPKDKDGEWDKI-TCPWNSAPGRTVRLFCPEYIIDF--DHTGH-----ALRHCS	:	77
AdoDHR	:	QRQRACEALMEGDAP-----DAAPDAPLRCAIAWDGV-NCWPEPTAGALAVQPCFDELNGIRYDIRQN-----ATRMCH	:	105
DmeCG8422	:	SVELQCLVQEHIEASTYGN-----DSGHCLTQFDSI-LCWPRTAGTTLAVLQCMDELQGIHYDSSKN-----ATRFCH	:	96
DmeCG13758	:	GVARVAQLALEPTVMDALPDSDTEQVLGLYCWNWTFDL-LCWPPPTAGVLAARMNCPGGFHGVDRKFAIR-----KCE	:	152
DmeCG32843	:	ALRAACETRLNASGQLAGSGG-PGAEAGTHCAGTFDEW-LCWPDPTAVGTSAYELCPDFITGFDPARYAH-----KECG	:	109
AgaP14164	:	EGLLALAMNASGAERCRLQQQAELLQDVACPSFFDMV-SCWPRTPPGTLAVLFCFAELKGVQYDSSQN-----ATRFCH	:	116
CelC13B9	:	ESYPLCKKMLDHNLFPER---DPNDTRIWCNATYDTV-LCWPPPTANSSVTLOCPHMKGLDP-NK-----NITKCH	:	114
CelC18B12	:	AEYAAQCQLKLE-----NSSYNPGGCSVDFFKS-LCWASAHIGQOMTRDCPFTFCTAIPGCEEIKDRYMVSRNCN	:	76
CbrCAE6326	:	TEYEACEKLKAQ-----NNSYTGSGCSVDFFKS-LCWASARLGOQOMTRDCPFFKCTSIPEGCEIKEYMVSRCN	:	75
CbrCAE7012	:	ESYPLCKKMLDHNLFPER---DPNDTRIWCNATYDTV-LCWPPPTANSSVTLOCPHMKGLDP-NKNVVKRCENITKCH	:	122

TM1				
HsaCALR	:	EKGWVFKHPENNRTWS-----NYTMCNAFTPEKLKNAYVLYLAIVGHSLSIFTLVISLGIFFVFR	:	174
HsaGLP1R	:	AEGWLQKDNSSLPWR-----DLSECESESKRGERSSPEEQLLFLYLIYTVGYALSFSALVIASAILLGFR	:	170
HsaCRFR1	:	ANGSWAARVNYSECQ-----EILNEEKSKVHYHVAVIYNLGHCSISLVALVAFVFLRLR	:	145
CinS5A	:	FDGSMVVRNDTGO-----SVQWVN-----RIQCDIIDIVTKIYTAGYIFSLVCMVIALFILMFYK	:	125
CinS5B	:	FDGSMVVRNDTGT-----SVQWVN-----RIQCDIDFVVIKIYTAGYIFSLVCMVIALFILMFYK	:	124
CinS93	:	ANGWFRYPPTQYQWT-----NYSRCQLQKQEAHVWAKYLAISGYSVSMITLLVALFIFFHFR	:	134
CinS372	:	RDGRWAMVRDNTNR-----TFSDYDSCNIPREEDAKMKSFIQLITVGYSFSLVALVFAMIIAYFK	:	137
CinS752	:	RDGRWAMVRDNTNR-----TFSDYSSCNIPREEDATMKSFIQLITVGYSFSLVALVFAMIIAYFK	:	137
AdoDHR	:	SNGTWRNYSYDVHCR-----ELVEAESDEDAAMAFVFFVGFCLSLVAIAVAIWIWIFYK	:	159
DmeCG8422	:	ANGTWEKYTNIDACALHP-----APESVPEFEVIVELPTLIYYIGYTLVSLVSLALIVFAYFK	:	155
DmeCG13758	:	LDGRWGSRRNATEVNPPGWTDYG-PCYKPEIIRLMQMGSKDFDAYIDIARRTRLEIVGLCLSLFALIVSLLIFCTFR	:	230
DmeCG32843	:	LDGEWFKHPLTNKTWS-----NYTTCVNELDLNRHTVNLISEVGYGTSLLAILLSLAILGYFK	:	168
AgaP14164	:	VDGTWNYTDYDRCEHLE-----QPPLPSPFEPEIELPTLIYFVGYSISLALVLAVALVIFYK	:	174
CelC13B9	:	MSGVWSGRNAGMGFTLPGWNTFTMCYTDEVIYIMQNLNNESTIAQEVARNARKLEFVGLGLSLVSLILAIISFYSFR	:	193
CelC18B12	:	SMGVWDDSNYTMCIK-----VVEEYAQCLQGFRCVCPDVLRLDVI SVSLTSLFVSVILLVAIIVLSIFD	:	141
CbrCAE6326	:	NQGVWDDTNYTMCIK-----VAEEYSQCLSGFCRVCPDFLRDMVI SVSLTSLIVSVILLVAIILFVSVFD	:	140
CbrCAE7012	:	MSGVWSGRNAAEMGFTLPGWNTFTMCYTDEVIYIMQNLNNESTIAQEVARNARKLEFVGLGLSLVSLILAIISFYSFR	:	201

**Figure 3**

**Comparison of the N-terminal end of metazoa family 2 GPCRs.** Multiple sequence alignment of the N-terminal domain of the family 2 GPCR receptor genes identified and characterised in protostomes and deuterostomes. Conserved cysteine residues are indicated by "\*" and the conserved amino acid motifs are boxed. The TM1 domain is annotated. Accession numbers of the human family 2 GPCRs: HsaGLP1R (P43220), HsaCRF1R (P34998) and HsaCALR (P30988). Accession numbers of the protostome family 2 receptor genes: House cricket (*Acheta domesticus*) Diuretic hormone receptor (AdoDHR, Q16983), *Drosophila melanogaster* (DmeCG13758, NP\_570007; DmeCG8422, NP\_610960; DmeCG32843, XP\_396046), *Anopheles gambiae* (AgaP14164, EAAI1768), *Caenorhabditis elegans* (CelC18B12.2, NP\_510496; CelC13B9.4, NP\_498465) and *Caenorhabditis briggsae* (CbrCAE70126, CAE70126; CbrCAE63268, CAE63268). EST data was used to obtain the N-terminal region of the incomplete receptor sequences (Additional file 1). The N-terminal region of *Ciona* CinS752 was predicted by NIX and that of CinS5A by sequence comparison with CinS5B. Only the clones for which a putative N-terminal domain was identified were included in the analysis. For figure simplicity, an arrow indicates the region of the receptor CbrCAE70126 that was eliminated since it did not align with any other sequences present.

**Figure 4**

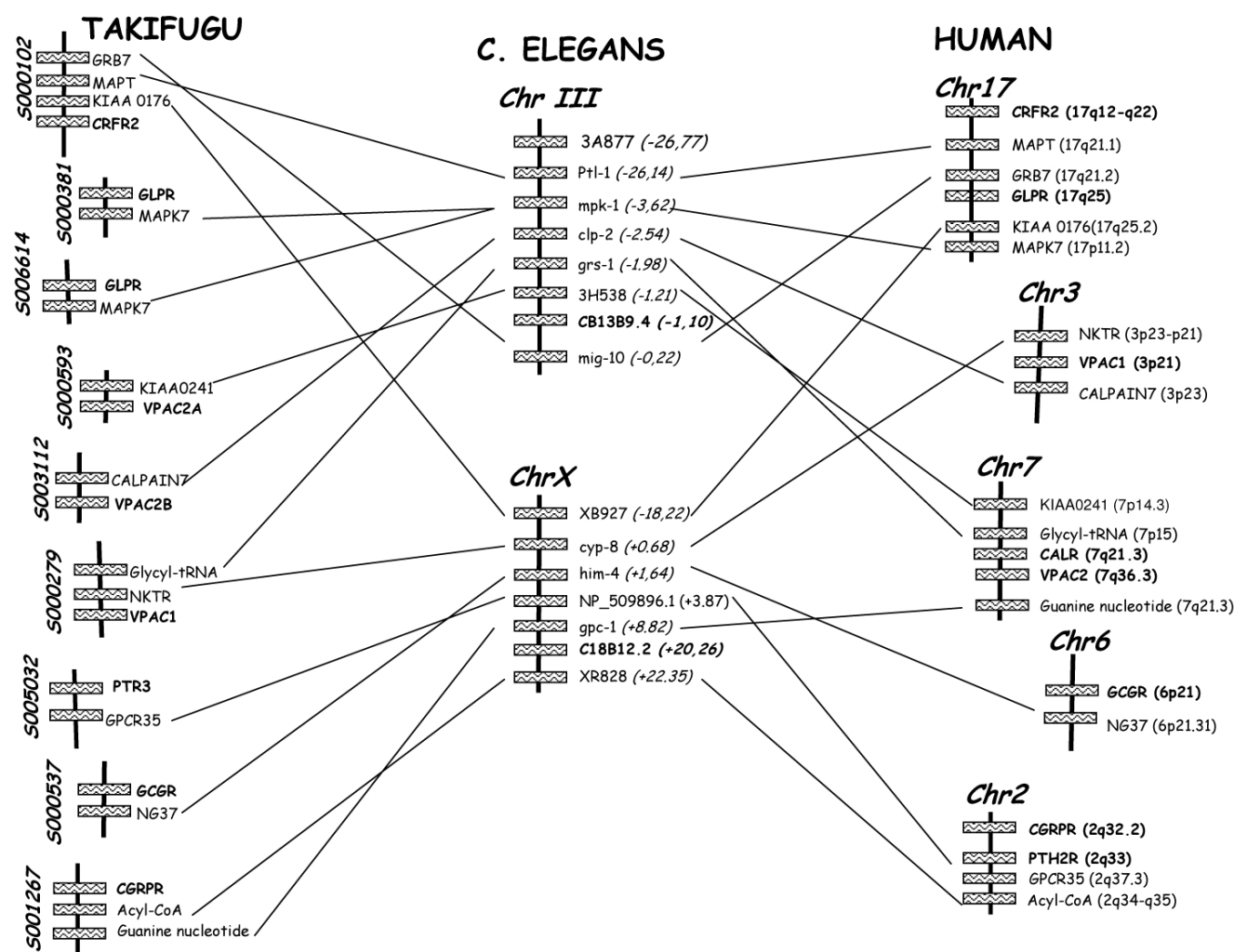
**Gene environment comparison between *C. elegans* and *Drosophila* family 2 GPCRs genomic regions.** Short-range linkage analysis of the region surrounding family 2 GPCRs on the *C. elegans* chromosome III and X with the *Drosophila* chromosome regions containing family 2 GPCR receptor genes. Genes are represented by horizontal bars and gene identification and chromosome position is given at the side. Family 2 GPCRs members are highlighted in bold. The lines represent the correspondence between the genes in each species. The dashed lines represent the common genes that were identified in both protostome and deuterostome genomes. For simplicity, only genes that are in common are represented.

In contrast with protostomes, two main gene structures were characterised in deuterostomes with TM4, TM5 and TM7 domains shared between two exons. In human and *Takifugu*, the CAL and CGRP receptors genes form a separate group with a different gene organisation from the other receptor genes. These receptors seem to have lost an intron between the exons that contain part of TM5 and TM6 and part of TM7 respectively and are composed of 7 exons with identical intron phases. The *Ciona* CinS5A, CinS5B, CinS50 and CinS273 receptor genes are also shorter and composed by 6 exons. With the exception of the CAL and CGRP receptors all human and *Takifugu* family 2 genes and *Ciona* CinS70, CinS93, CinS372 and CinS752 receptors are composed of 8 exons and share identical intron phases.

#### Expression analysis

RT-PCR with specific primers for each receptor gene was carried out using total RNA extracted from whole adult *C.*

*elegans*, *Drosophila*, mosquito and *Ciona* (results not shown). The nematode receptor genes CelC18B12.2 and CelC13B9.4, the *Drosophila* DmeCG13758, DmeCG32843 and DmeCG8422 were successfully amplified. The tissue distribution was refined in *Ciona* with expression analyses carried out in the intestine, pharyngeal basket, gonads, endostyle and cerebral ganglion. The putative family 2 GPCR encoded by CinS70, was expressed in all tissues analysed, but was mainly present in the intestine, gonads and cerebral ganglion. The CinS5A gene expression had a similar tissue distribution to CinS70. However its duplicate, CinS5B had a more limited distribution and was only expressed in the intestine. The receptors CinS273 and CinS752 were only expressed in gonad and cerebral ganglion tissue, whilst CinS93 was expressed in the endostyle. CinS50 was restricted to the ovary where it was weakly expressed. It was not possible to amplify AgENSANGP00000014164 and CinS372 receptors.

**Figure 5**

**Gene environment comparison between *C. elegans* and *Takifugu* and Human family 2 GPCRs genomic regions.** Short-range linkage analysis of the region surrounding family 2 GPCRs on the *C. elegans* chromosome III and *Takifugu* and human chromosome regions containing family 2 GPCR receptor genes. Genes are represented by horizontal bars and gene identification and chromosome position is given at the side. Family 2 GPCRs members are highlighted in bold. The lines represent the correspondence between the genes in each species. The dashed lines represent the common genes that were identified in both protostome and deuterostome genomes. For simplicity, only genes that are in common are represented.

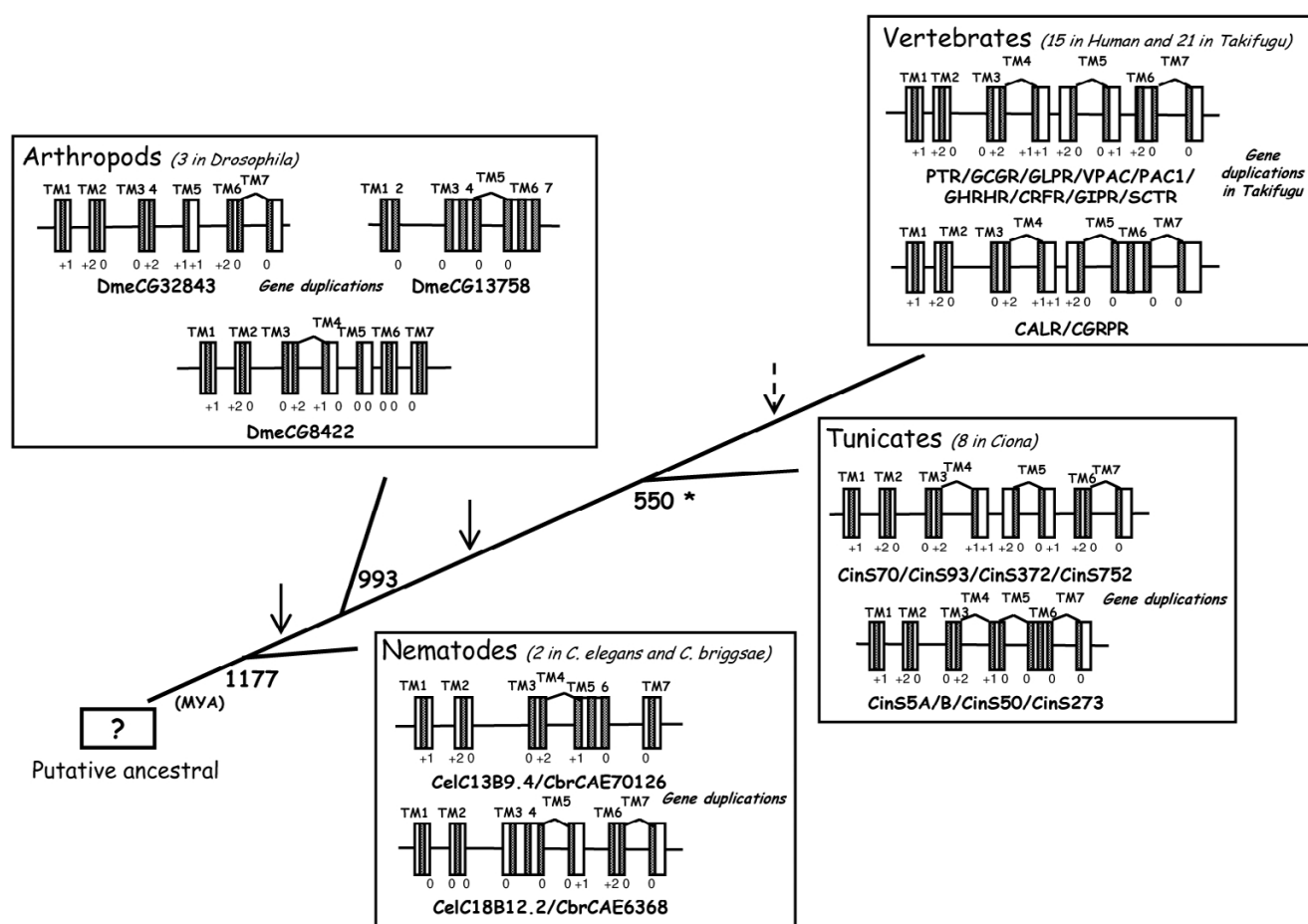
## Discussion

In total 16 putative family 2 GPCR members were identified and characterised in the protostome and in the tunicate genome using the human and *Takifugu* TM domain sequences. In the human genome a total of 50 family 2 GPCR receptors have been described [6,11,12] although ligands have only been assigned for 15 family 2 GPCR members [16,26]. The evolution of this latter receptor group is the major subject of this study.

To avoid the inclusion of potential intronic regions which could bias analysis, the invertebrate genes were manually edited using expression data available and sequence simi-

larity for the homologue genes and the *in silico* analysis performed was restricted for receptor conserved motifs at N-terminal and TM domains. The putative invertebrate receptors identified share the general characteristics of family 2 GPCR members with the presence of conserved cysteine residues [27,28] and several highly conserved amino acids and protein motifs at the N-terminal domain which is involved in ligand binding [22]. Analysis of unicellular organism genomes such as the prokaryote *E. coli* and the eukaryote *S. cerevisiae* failed to reveal any putative family 2 receptor genes suggesting that these receptor genes are characteristic of metazoan genomes.



**Figure 6**

**Evolution of family 2 GPCRs in the metazoa.** Diagram illustrating the increase in gene number and gene structure complexity of the TM domains regions of family 2 GPCR receptor genes in protostomes (nematodes and arthropods) and deuterostomes (tunicate and vertebrate). The proposed species-specific gene duplications are indicated in the figure. Putative gene or genome duplication events of ancestral family 2 GPCRs during metazoan evolution are indicated by arrows (dashed arrow – represent several duplication events within the chordate lineage). The number of receptors identified in each species is within brackets. Exons are represented by blocks and introns by lines and the molecular evolutionary time for each species is indicated in MYA (million years ago) and was obtained from Hedges and Kumar, 2003 [82] and (\*) from Dehal et al. 2002 [42]. The TM domain regions are represented by shaded regions and numbered. Intron/exon boundary phases are indicated below each exon. The tunicate and vertebrates family 2 GPCR members have been grouped according to their common gene organisation. The figure is not drawn to scale.

The protostome genes identified are more like the vertebrate CAL/CGRP and CRF receptors subfamilies suggesting that ancestral family 2 GPCR members were most like these genes. In vertebrates these receptors are associated with calcium homeostasis and the stress axis respectively, and appropriate functioning is essential for survival [29-32]. No studies are available describing the role of these receptors in protostomes and their classification was based upon their sequence similarity, so they may not be functional orthologues. Recently, expression studies carried out using the *Drosophila* DmeCG8244 and DmeCG32843 (equivalent to CG17415) receptor genes,

which are very similar in sequence to the vertebrate CRF and CAL/CGRP receptors, respectively, were indeed found to be functional orthologues. These studies revealed that DmeCG8244 and DmCG32843 in the presence of insect diuretic hormones (DH) were functional and activate a similar intracellular signalling pathway to the vertebrate receptor genes [33,34].

It is generally accepted that complexity of vertebrates and the origin and development of physiological systems is a consequence of the acquisition of new genes by gene or exon duplications that occurred during chordate evolu-

tion ([10,35-38]. The absence of sequence homologues of the other members of family 2 in protostomes may be a consequence of the relatively low complexity of nematodes and arthropods when compared with vertebrates. For example, it is known that VIP, PACAP, SCT, GCG, GLPs and GIP peptides and their receptors are mainly associated with the nervous and gastrointestinal systems. In protostomes these biological structures are of very low complexity [39] and an organised gastroenteropancreatic system has only been identified after the divergence of tunicates. In general, a similar evolutionary profile occurs during the development of metazoan nervous systems (with the exception of cephalopod molluscs, which also have a highly developed nervous system). The occurrence of an organised brain and complex central nervous structure is only present in vertebrates. In the majority of protostomes and early deuterostomes the nervous system is mainly characterised by the presence of simple structures such as cerebral and head ganglions to which several nerve networks are connected and like the gastroenteropancreatic system it has been evolving by varying degrees of complexity throughout metazoan evolution [39].

The existence of an increasing gene number of family 2 GPCR members in the metazoan lineage clearly suggests that evolution of this gene family results from a series of duplications. Several gene duplication events have been proposed to have occurred in chordate genomes. However if they are a consequence of two total genome duplication events (2R theory, [10,40-42]) or a result of independent single gene duplications ([43,44]) is still under debate. The 2R theory has been generally accepted to justify the presence of gene family members and novel genes in higher vertebrate genomes when compared with early chordate and invertebrate species. Family 2 GPCR members have been identified in the majority of vertebrate genomes [17,20] and searches carried out on the amphibian and chicken genomes identified an equivalent number of gene family members (data not shown) to that found in the human genome. In particular putative SCT receptor genes, lacking in teleost genomes, were identified as previously described by Langerstrom et al [46]. The absence of an equivalent SCT receptor sequence in teleosts but its identification in amphibian and avian genomes may indicate that this receptor either, i) arose after the divergence of the fishes or, ii) was lost in the fish lineage and studies on ancient fishes (eg. Agnatha) should help to clarify this issue.

The protostome receptors contain the most divergent gene structures when compared with those of deuterostomes (intron number, TM domain distribution and intron phases). This is probably a consequence of the higher rate of chromosomal and gene rearrangements when compared to early chordates and vertebrates [47-50]. The rea-

son behind differences in intron numbers between protostomes and deuterostomes remains to be established. The precursor gene in Urbilateria (last common ancestors of protostomes and deuterostomes) remains to be identified and therefore the difference in intron numbers can be explained either by intron gain during metazoan evolution after the divergence of the protostome and deuterostome [51-53] or by intron loss in the protostome lineage [54,55]. Interestingly, the vertebrate CAL and CGRP receptor gene organisation is more like the tunicate receptors than other vertebrate family 2 members and the way in which these receptors function is also different. Accessory single transmembrane proteins (Receptor activity modifying proteins, RAMPs [56,57]) interact with both CAL and CGRP receptors and alter their affinity profile for the ligands (CAL, CGRP, adrenomedullin and amylin) [58,59]. In fact CAL receptor-RAMP heterodimerization is essential for receptor function but not for the other family 2 members [58]. It remains to be established if such functional constraints can influence gene evolution and in particular CAL/CGRP receptors gene evolution.

Expression analyses indicates that with the exception of the single mosquito receptor and CinS372, all protostome and tunicate genes are expressed. In general, the tissue distribution of the *Ciona* receptors mirrors the expression of the vertebrate family 2 GPCRs sequence homologues. For example, the duplicate CinS5A and CinS5B receptors and CinS70 which share sequence similarity for the vertebrate brain-gut peptide GCG, GLP, GIP receptors were found to be expressed in *Ciona* intestine, gonads and neural ganglion suggesting that like the vertebrate homologues they may also have a role in the gastrointestinal, reproductive and nervous systems [60,61]. The function of the vertebrate receptors in the reproductive system is not clear, but in the nervous and gastrointestinal systems they are involved in carbohydrate, amino acid and lipid metabolism [60,62]. It remains to be established if the tunicate receptors localised in these tissue have a similar functional role.

The *Ciona* CAL/CGRP (CinS93) homologue was only expressed in the endostyle whilst the tunicate PTR-like receptor homologue (CinS752) was present in the neural ganglion and the gonads. In vertebrates these receptors are found to have an important role in the endocrine regulation of calcium mediated by CAL and PTH hormones [63]. The presence in *Ciona* of the vertebrate homologue receptors may suggest that elements of calcium homeostasis are conserved between tunicates and vertebrates. Moreover, the expression of CinS93 in *Ciona* endostyle, the homologue of the vertebrate thyroid gland, the site of CAL production [64], further supports this hypothesis. The CinS50 and CinS273 are the sequence homologues of the vertebrate CRF receptors, which are mainly associated

with stress response [65]. In vertebrates both receptors are found to be expressed in nervous tissue and CRF1 receptor was also found in the gonads and CRF2 receptor in the gastrointestinal tract [66,67]. The pattern of expression of the tunicate CinS50 and CinS273 was similar and both receptors were expressed in the gonads whilst CinS273 was further detected in the neural ganglion suggesting they may also play a functional role in the nervous and reproductive systems in tunicates. Functional and ligand-binding studies are required to characterise the physiological role of the tunicate receptors. Moreover, the isolation and characterisation of their putative ligand peptides which have yet to be comprehensively described will be essential to understanding of their function.

The evolution of family 2 GPCR receptor genes in proto-stomes and deuterostomes is probably the result of a combination of species-specific gene duplications and gene or genome duplication events in ancestral gene precursors (Figure 6). For example, the *Drosophila* DmeCG8422 and DmeCG32843 map to chromosome 2R and share a similar gene organisation. This suggests that they arose by a specific gene duplication event. Based on their sequence similarity, gene organisation and intron phases, the *Drosophila* DmeCG8422 and DmeCG32843 receptor genes appear to be the orthologues of the nematode CelC13B9.4 and *Ciona* CinS50/CinS273 and CinS93 receptors, respectively. In tunicates, 3 different family 2 GPCR ancestral gene precursors probably existed: the gene precursor for CinS50/CinS273, the gene precursor for CinS93 and a common gene precursor for CinS70/CinS5A/CinS5B/CinS372/CinS752. These are proposed to be the origin of the vertebrate CRF, CAL/CGRP genes and remaining deuterostome family 2 GPCRs, respectively.

## Conclusion

Putative family 2 receptor genes were isolated and characterised from a number of different invertebrate genomes. This study provides for the first time a comprehensive description of the gene sequence, structure and expression of family 2 GPCRs members in invertebrates providing important clues about their origin and evolution along

the metazoan divergence. The CAL/CGRP and CRF receptors are proposed to be the first family 2 members to evolve in contrast to SCT receptors which seem to have evolved much later and are only present in tetrapods. Studies such as this, can via a mixture of *in vitro* and *in silico* approaches, contribute to a better understanding of gene regulation in vertebrates.

## Methods

### Sequence database searches

Sequence database searches were carried out on the genomes of the prokaryote *Escherichia coli* (*E. coli*), the unicellular eukaryote *Saccharomyces cerevisiae* (*S. cerevisiae*), the nematodes *Caenorhabditis elegans* (*C. elegans*) and *Caenorhabditis briggsae* (*C. briggsae*), the insects, *Drosophila melanogaster* (*D. melanogaster*) and mosquito *Anopheles gambiae* (*A. gambiae*) and in the tunicate *Ciona intestinalis* (*C. intestinalis*) (Table 2) using the TM domains of the 15 human [12] and the 21 *Takifugu rubripes* from Cardoso et al., 2005[17] family 2 receptor genes. The TM domains of the vertebrate family 2 GPCRs were concatenated and used in conjunction with the BLASTP and TBLASTN algorithms [68] to interrogate invertebrate genomes. The invertebrate *in silico* predicted sequences were identified based on their sequence similarity for the vertebrate receptor genes using a cut-off *E* value higher than 10 and their sequences were manually edited according to their similarity for the homologue genes in vertebrates and EST data available (Additional file 1). The identity of the invertebrate genes was further confirmed against the GPCR database at CMBI [69] and in order to substantiate previous searches (identify putative receptors that were not identified) they were further used to search all the genome databases used in this analysis (Table 2).

### Gene organisation of invertebrate family 2 GPCRs

The gene organisation of the invertebrate family 2 members was manually characterised. This approach was complemented using the available protostome and *Ciona* EST data to identify putative N and C terminal ends of the protein (Additional file 1). The presence of TM domain regions was verified using the TMHMM Server v. 2.0 [70] and their positions were subsequently confirmed by mul-

**Table 2: Databases used to identify putative family 2 GPCRs in invertebrate genomes**

Organism	Database
<i>Escherichia coli</i>	<a href="http://www.ncbi.nih.gov">http://www.ncbi.nih.gov</a>
<i>Saccharomyces cerevisiae</i>	<a href="http://www.yeastgenome.org/">http://www.yeastgenome.org/</a>
<i>Caenorhabditis elegans</i> and <i>briggsae</i>	<a href="http://www.sanger.ac.uk">http://www.sanger.ac.uk</a>
<i>Drosophila melanogaster</i>	<a href="http://www.wormbase.org/">http://www.wormbase.org/</a>
	<a href="http://flybase.bio.indiana.edu/">http://flybase.bio.indiana.edu/</a>
	<a href="http://www.ensembl.org">http://www.ensembl.org</a>
<i>Anopheles gambiae</i>	<a href="http://www.ensembl.org">http://www.ensembl.org</a>
<i>Ciona intestinalis</i>	<a href="http://genome.jgi-psf.org">http://genome.jgi-psf.org</a>

multiple sequence comparison alignments with the vertebrate homologues based on PRINTS annotation [71]. The gene structures of the human receptors were characterised using the Spidey mRNA-to-genomic alignment programme [72] and the *Takifugu* receptor gene organisation characterised as described in Cardoso et al, 2005[17].

### Linkage analysis

The gene environments of the protostome (*C. elegans* and *Drosophila*) and deuterostome (human and *Takifugu*) receptor genes were compared using a sequence similarity approach. The human, *C. elegans* and *Drosophila* gene environments were accessed using the NCBI Mapview interfaces [73]. The gene environment of the *Takifugu* scaffolds (release17/05) was accessed using NIX annotation [74] and the neighbouring genes were used to search for orthologues in human, *C. elegans* and *Drosophila* genomes using the TBLASTX algorithm [75].

### Sequence comparison and phylogenetic analysis

Sequence alignments of the predicted protostome and deuterostome receptor protein sequences were carried out using the Clustal X programme [76] (Blosum matrix, Gap opening penalty 10, Gap extension 0.2) with and percentage similarity were calculated using GeneDoc [77]. The evolutionary analysis between the protostome and deuterostome receptor genes was carried out using the TM domains that were complete and common to all receptor genes (TM2, TM4, TM5 and TM6) following a similar strategy has previously described [17]. Manual editing of the *Takifugu* family 2 GPCRs did not identify TM1 domain of TruS012367, the TM5 of TruCRFR2 was found to be incomplete and TM3 of TruS000381 was frameshifted.

The four TM domain sequences common to all metazoan were concatenated and aligned using the ClustalX programme as described. The alignment produced (length 98, with 99 informative sites) was used for phylogenetic analysis using the neighbour joining, maximum parsimony and minimum evolution methods with 1000 bootstrap replicates in the MEGA 3.1 phylogenetic programme [78]. Multiple sequence alignments were also carried out using the manually edited protostome and deuterostome receptors within each family 2 GPCR group using the ClustalX programme (according to the parameters previously described) in order to further identify conserved protein motifs or amino acid residues at the N-terminal regions that might be involved in ligand-binding.

### Expression analysis

In order to investigate the expression of the putative protostome and tunicate receptors RT-PCR was carried out using cDNA produced from whole individual organisms. Total RNA from adult individuals was extracted from the nematodes, *Drosophila* and *Ciona* with TRI reagent (Sigma-

Aldrich, Spain) according to the manufactures instructions. 1–2 µg of total RNA was used for cDNA synthesis and each reaction was performed as follows, 1xRT-PCR buffer (Invitrogen), 0.25 mM dNTPs (Amersham-Biosciences, UK), 0.05 µg/µl random hexameric oligonucleotides, 1 U MMLV-RT (200 U/µl) (Promega, USA) and 0.2 U RNAGuard 36.3 U/µl (Amersham-Biosciences). Specific primers for each receptor gene were designed spanning different exons to detect potential genomic contamination. A control PCR using primers for housekeeping genes was also performed in order to control the amount of cDNA utilised in each reaction. Specific primers for *Ciona* 18 S ribosomal protein were designed but sea bream 18 S and β-actin primers were routinely used in the *Drosophila* and the nematode. All primers (including housekeeping control sequences) used are described in Table 3. All the PCR reactions were performed with 1xPCR buffer (Euroclone, Italy), 1.5 mM MgCl<sub>2</sub> (Euroclone), 0.2 mM dNTPs (Amersham-Biosciences), 1 mM of each primer (Forward and Reverse) EuroTaq DNA Polymerase 5 U/µl (Euroclone) and DNase Free water (Sigma-Aldrich) for a 25 µl final reaction volume. Amplification of all the genes was carried out using a standard cycle with an initial denaturing step of 93°C for 2 minutes, followed by 35 cycles of: 30 s at 93°C, the annealing temperatures of primers for 60 s and 72°C for 30 seconds followed by a final chain extension step of 72°C for 5 minutes. The reaction products were cloned into pGEMT-easy vector (Promega) and sequenced to confirm their identity.

### Abbreviations

GPCRs (G-protein coupled receptors); TM (transmembrane); CAL (Calcitonin), CGRP (Calcitonin Gene Related Peptide); CRF (Corticotrophin Releasing Factor); PTH (Parathyroid Hormone); VIP (Vasoactive Intestinal Polypeptide); PACAP (Pituitary Adenylate-Cyclase Activating Polypeptide); SCT (Secretin); GCG (Glucagon); GLP (Glucagon Like Peptide); GIP (Glucose Insulinotropic Peptide); VPAC (Vasoactive Intestinal Polypeptide receptor); PAC1 (Pituitary Adenylate- Cyclase Activating Polypeptide receptor); PTR (Parathyroid Hormone receptor)

### Authors' contributions

The majority of the work here described was carried by JCRC in collaboration with FAV and VCP. MSC and DMP planned the study, and critically revised the manuscript for important intellectual content and data analysis. All authors read and approved the final manuscript

**Table 3: Primer pairs used to amplify by RT-PCR protostome and tunicate family 2 GPCRs**

Primer pairs	
<b>Nematode</b> <i>C. elegans</i> (Cel)	CelC13B9.4F1 5' -gatacacgaatttggtgtaatgcc-3'/CelC13B9.4R1 5' -gttcgtgtgaggaccattttcac-3' CelC18B12.2F1 5' -ccattcacattttgactgcaatt-3'/CelC18B12.2R1 5' -gaaccagagaatagctttgcaaat-3'
<b>Insect</b> <i>D. melanogaste</i> (Dme) <i>A. gambiae</i> (Aga)	DmeCG13758F 5' -gagattatccgtctcatgca-3'/DmeCG13758R 5' -cgcggttcaactgtggcgtt-3' DmeCG8422F 5' -agctgccaccattatctac-3'/DmeCG8422R 5' -gttctgggtaagctgaatggt-3' DmeCG32843F 5' -gcatacagctgcacatgaat-3'/DmeCG32843R 5' -cgccgagatgatttcgtatg-3' AgaPI4164F 5' -agcttcgagccggaaattgag-3'/AgaPI4164R 5' -ttcgtgatcagcaccacatgat-3'
<b>Tunicate</b> <i>C. intestinalis</i> (Cin)	CinS93F 5' -taccgcttggtgatttctg-3'/CinS93R 5' -aacgacatagagtagcaacga-3' CinS372F 5' -aacgacatagagtagcaacga-3'/CinS372R 5' -cgagcagaaaatcaaatatg-3' CinS752F 5' -attgctcacgtgacggtaga-3'/CinS752R 5' -ctccacaaaatttctgttc-3' CinS273F 5' -gttttagaaacggaaacatg-3'/CinS273R 5' -agaaaaactgttccggtt-3' CinS50F 5' -gatccataggaagtgaaag-3'/CinS50R 5' -aatgaaataatttctcgggt-3' CinS70F 5' -cagctgtcgtactagtaataac-3'/CinS70R 5' -gtatacagagacgatttccctg-3' CinS5AF 5' -cctggattttgatgggtat-3'/CinS5AR 5' -attaggtatcatattgtttac-3' CinS5BF 5' -cctggatttcttgatagacaa-3'/CinS5BR 5' -agttcgtaaagcgttgtagat-3'
<b>Ciona 18S control</b>	Cin18SFwd 5' -cggagaagtttcagcaca-3'/Cin18SRev 5' - agtgcgcaaacccctgt-3'
<b>18S control</b>	Sb18SFwd 5' -tcaagaacgaaagtcggagg-3'/Sb18SRev 5' -ggacatctaagggcatcaca-3'
<b>β-actin control</b>	SbDebactF3 5' -ggccgcgacctcacagactac-3'/SbDebactR2 5' -accgaggaaggatggctggaa-3'

## Additional material

### Additional File 1

**List of the protostome (nematodes and arthropods) and tunicate (Ciona) putative family 2 GPCRs identified.** The total size of each receptor protein sequence used in the in silico analysis performed is indicated within brackets and their sequences are available as additional data (Additional files 3, 4, 5, 6). The EST data available for each receptor and source of information are also indicated. No ESTs were available for the *C. briggsae* and mosquito receptor genes identified. The N-terminal region of CinS5A was identified by sequence comparison with the paralogous gene CinS5B and the C-terminal end of CinS93 was predicted using exon prediction programmes and sequence similarity approaches with the vertebrate homologue genes.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2148-6-108-S1.pdf>]

### Additional File 2

**Sequences of the protostome and deuterostome family 2 GPCRs TM domains used in phylogenetic analysis.** Amino acid sequences of the human (*Hsa*), Takifugu (*Tru*), Ciona (*Cin*), Drosophila (*Dme*), Mosquito (*Aga*), *C. elegans* (*Cel*) and *C. briggsae* (*Cbr*) TM2, TM4, TM5 and TM6 domain regions used in the construction of the phylogenetic tree.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2148-6-108-S2.pdf>]

### Additional File 3

**Multiple sequence comparisons of the metazoan CALR/CGRPR.** Multiple sequence alignment carried out with the protostome and deuterostome putative CALR/CGRPR protein sequences. The protostome, tunicate and Takifugu receptor sequences were manually edited having in consideration their sequence similarity, identification of splice sites consensus sequences (AG/GT) and the existence of EST data. Conserved cysteine residues are indicated by (●) and TM domains named. The N-terminal regions were annotated according to their level of conservation of the protostome and deuterostome receptors. The amino acid residues annotated with closed boxes have been previously identified in Figure 3 and the novel amino acid residues and protein motifs are annotated by open boxes. Incomplete sequences are due to gaps or low quality sequence data within receptor gene sequences. The Takifugu S012367 receptor sequence was not included since TM1 was not identified. The existence of putative intronic sequences was when possible investigated using the EST data available (Additional file 1). The start codon was chosen as the methionine in the correct frame of the first exon and the end of each receptor gene was chosen as the first stop codon in the correct frame of the last exon.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2148-6-108-S3.pdf>]

### Additional File 4

**Multiple sequence comparisons of the metazoan CRFR.** Multiple sequence alignment carried out with the protostome and deuterostome putative CRFR protein sequences. The protostome, tunicate and Takifugu receptor sequences were manually edited having in consideration their sequence similarity, identification of splice sites consensus sequences (AG/GT) and the existence of EST data. Conserved cysteine residues are indicated by (●) and TM domains named. The N-terminal regions were annotated according to their level of conservation of the protostome and deuterostome receptors. The amino acid residues annotated with closed boxes have been previously identified in Figure 3 and the novel amino acid residues and protein motifs are annotated by open boxes. Incomplete sequences are due to gaps or low quality sequence data within receptor gene sequences. TruCRFR2, CinS273 and AgaP14164 have incomplete N-terminal ends and TruCRFR2 was also found to contain an incomplete intracellular loop 3. Despite the availability of EST data for N-terminal of CinS50 two potential methionines were identified and for these reason the putative exon 1 sequence was not included. The existence of putative intronic sequences was when possible investigated using the EST data available (Additional file 1). The start codon was chosen as the methionine in the correct frame of the first exon and the end of each receptor gene was chosen as the first stop codon in the correct frame of the last exon.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2148-6-108-S4.pdf>]

### Additional File 5

**Multiple sequence comparisons of the metazoan PTR.** Multiple sequence alignment carried out with the putative PTR deuterostome protein sequences. Tunicate and Takifugu receptor sequences were manually edited having in consideration their sequence similarity, identification of splice sites consensus sequences (AG/GT) and the existence of EST data. Conserved cysteine residues are indicated by (●) and TM domains named. The N-terminal regions were annotated according to their level of conservation. The amino acid residues annotated with closed boxes have been previously identified in Figure 3 and the novel amino acid residues and protein motifs are annotated by open boxes. Incomplete sequences are due to gaps or low quality sequence data within receptor gene sequences. The existence of putative intronic sequences was when possible investigated using the EST data available (Additional file 1). The start codon was chosen as the methionine in the correct frame of the first exon and the end of each receptor gene was chosen as the first stop codon in the correct frame of the last exon.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2148-6-108-S5.pdf>]

### Additional File 6

**Multiple sequence comparisons of the metazoan CGCR/GLPR/GIPR.** Multiple sequence alignment carried out with the deuterostome putative CGCR/GLPR/GIPR protein sequences. Tunicate and Takifugu sequences were manually edited having in consideration their sequence similarity, identification of splice sites consensus sequences (AG/GT) and the existence of EST data. Conserved cysteine residues are indicated by (●) and TM domains named. The N-terminal regions were annotated according to their level of conservation. The amino acid residues annotated with closed boxes have been previously identified in Figure 3 and the novel amino acid residues and protein motifs are annotated by open boxes.

Incomplete sequences are due to gaps or low quality sequence data within receptor gene sequences. Takifugu S006614 and S007267 and CinS70 have incomplete N-terminal regions and the Takifugu receptor S000381 was not included in the alignment since TM3 is frameshifted. The existence of putative intronic sequences was when possible investigated using the EST data available (Additional file 1). The start codon was chosen as the methionine in the correct frame of the first exon and the end of each receptor gene was chosen as the first stop codon in the correct frame of the last exon. The Ciona CinS5A, CinS5B and CinS70 were only compared with the vertebrate CGCR/GLPR/GIPR with which they share higher sequence similarity

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2148-6-108-S6.pdf>]

### Acknowledgements

This work was supported by a Fundação para a Ciência e Tecnologia/Centro de Ciências do Mar pluriannual grant. JC was funded by FCT grant BPD/14449/03. Adult nematodes *C. elegans* were kindly supplied by Diogo Manoel and Dr. Henrique Teotónio from Instituto Gulbenkian para a Ciência, Portugal, the adult *Drosophila* by Dr. Dephine Depoix from the Institut de Recherche sur la Biologie de l'Insecte, France, and mosquito cDNAs were kindly supplied by Dr. Henrique Silveira from the Instituto de Medicina Tropical, Lisboa.

### References

1. Fredriksson R, Schiöth HB: **The repertoire of G-protein-coupled receptors in fully sequenced genomes.** *Mol Pharmacol* 2005, **67**(5):1414-1425.
2. Takeda S, Kadowaki S, Haga T, Takaesu H, Mitaku S: **Identification of G protein-coupled receptor genes from the human genome sequence.** *FEBS Lett* 2002, **520**(1-3):97-101.
3. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, et al.: **Initial sequencing and analysis of the human genome.** *Nature* 2001, **409**(6822):860-921.
4. Ulrich CD 2nd, Holtmann M, Miller LJ: **Secretin and vasoactive intestinal peptide receptors: members of a unique family of G protein-coupled receptors.** *Gastroenterology* 1998, **114**(2):382-397.
5. Ulloa-Aguirre A, Stanislaus D, Janovick JA, Conn PM: **Structure-activity relationships of G protein-coupled receptors.** *Arch Med Res* 1999, **30**(6):420-435.
6. Fredriksson R, Lagerström MC, Lundin LG, Schiöth HB: **The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints.** *Mol Pharmacol* 2003, **63**(6):1256-1272.
7. Graul RC, Sadee W: **Evolutionary relationships among G protein-coupled receptors using a clustered database approach.** *AAPS PharmSci* 2001, **3**(2):E12.
8. Haldane JBS: *The causes of evolution* New Jersey: Princeton University Press/1990; 1932.

9. Muller HJ: **The origination of chromatin deficiencies as minute deletions subject to insertion elsewhere.** *Genetica* 1935, **17**:237-252.
10. Ohno S: **Evolution by gene duplication.** Berlin, New York, Springer-Verlag; 1970.
11. Stacey M, Lin HH, Gordon S, McKnight AJ: **LNB-TM7, a group of seven-transmembrane proteins related to family-B G-protein-coupled receptors.** *Trends Biochem Sci* 2000, **25**(6):284-289.
12. The International Union of Pharmacology [<http://www.iuphar-db.org/list>]
13. Donnelly D: **The arrangement of the transmembrane helices in the secretin receptor family of G-protein-coupled receptors.** *FEBS Lett* 1997, **409**(3):431-436.
14. Harmar AJ: **Family-B G-protein-coupled receptors.** *Genome Biology* 2001, **2**(12):3012.3011-3013.3010.
15. Vaudry D, Gonzalez BJ, Basille M, Yon L, Fournier A, Vaudry H: **Pituitary adenylate cyclase-activating polypeptide and its receptors: from structure to functions.** *Pharmacol Rev* 2000, **52**(2):269-324.
16. Sherwood NM, Krueck SL, McRory JE: **The origin and function of the pituitary adenylate cyclase-activating polypeptide (PACAP)/glucagon superfamily.** *Endocr Rev* 2000, **21**(6):619-670.
17. Cardoso JC, Clark MS, Viera FA, Bridge PD, Gilles A, Power DM: **The secretin G-protein-coupled receptor family: teleost receptors.** *J Mol Endocrinol* 2005, **34**(3):753-765.
18. Metpally RP, Sowdhamini R: **Genome wide survey of G protein-coupled receptors in Tetraodon nigroviridis.** *BMC Evol Biol* 2005, **5**:41.
19. Hewes RS, Taghert PH: **Neuropeptides and neuropeptide receptors in the Drosophila melanogaster genome.** *Genome Res* 2001, **11**(6):1126-1142.
20. Cardoso JC, Power DM, Elgar G, Clark MS: **Duplicated receptors for VIP and PACAP (VPACIR and PACIR) in a teleost fish, Fugu rubripes.** *J Mol Endocrinol* 2004, **33**(2):411-428.
21. DeAlmeida VI, Mayo KE: **Identification of binding domains of the growth hormone-releasing hormone receptor by analysis of mutant and chimeric receptor proteins.** *Mol Endocrinol* 1998, **12**(5):750-765.
22. Laburthe M, Couvineau A, Marie JC: **VPAC receptors for VIP and PACAP.** *Receptors Channels* 2002, **8**(3-4):137-153.
23. Wille S, Sydow S, Palchaudhuri MR, Spiess J, Dautzenberg FM: **Identification of amino acids in the N-terminal domain of corticotropin-releasing factor receptor 1 that are important determinants of high-affinity ligand binding.** *J Neurochem* 1999, **72**(1):388-395.
24. Dautzenberg FM, Kilpatrick GJ, Wille S, Hauger RL: **The ligand-selective domains of corticotropin-releasing factor type 1 and type 2 receptor reside in different extracellular domains: generation of chimeric receptors with a novel ligand-selective profile.** *J Neurochem* 1999, **73**(2):821-829.
25. Gardella TJ, Juppner H: **Molecular properties of the PTH/PTHrP receptor.** *Trends Endocrinol Metab* 2001, **12**(5):210-217.
26. Foord SM, Bonner TI, Neubig RR, Rosser EM, Pin JP, Davenport AP, Spedding M, Harmar AJ: **International Union of Pharmacology. XLVI. G protein-coupled receptor list.** *Pharmacol Rev* 2005, **57**(2):279-288.
27. Segre GV, Goldring SR: **Receptors for secretin, parathyroid hormone (PTH)/PTH-related peptide, vasoactive intestinal peptide, glucagon-like peptide 1, growth hormone-releasing hormone, and glucagon belong to a newly discovered G-protein-linked receptor family.** *Trends Endocrinol Metab* 1993, **4**:309.
28. Laburthe M, Couvineau A, Gaudin P, Maoret JJ, Rouyer-Fessard C, Nicole P: **Receptors for VIP, PACAP, secretin, GRF, glucagon, GLP-1, and other members of their new family of G protein-linked receptors: structure-function relationship with special reference to the human VIP-1 receptor.** *Ann N Y Acad Sci* 1996, **805**:94-109. discussion 110-101
29. Carmeliet G, Van Cromphaut S, Daci E, Maes C, Bouillon R: **Disorders of calcium homeostasis.** *Best Pract Res Clin Endocrinol Metab* 2003, **17**(4):529-546.
30. Heinrichs SC, Menzaghi F, Merlo Pich E, Britton KT, Koob GF: **The role of CRF in behavioral aspects of stress.** *Ann N Y Acad Sci* 1995, **771**:92-104.
31. Carrasco GA, Van de Kar LD: **Neuroendocrine pharmacology of stress.** *Eur J Pharmacol* 2003, **463**(1-3):235-272.
32. Zaidi M, Moonga BS, Bevis PJ, Bascal ZA, Breimer LH: **The calcitonin gene peptides: biology and clinical relevance.** *Crit Rev Clin Lab Sci* 1990, **28**(2):109-174.
33. Johnson EC, Bohn LM, Taghert PH: **Drosophila CG8422 encodes a functional diuretic hormone receptor.** *J Exp Biol* 2004, **207**(Pt 5):743-748.
34. Johnson EC, Shafer OT, Trigg JS, Park J, Schooley DA, Dow JA, Taghert PH: **A novel diuretic hormone receptor in Drosophila: evidence for conservation of CGRP signaling.** *J Exp Biol* 2005, **208**(Pt 7):1239-1246.
35. Ohno S: **Gene duplication and the uniqueness of vertebrate genomes circa 1970-1999.** *Semin Cell Dev Biol* 1999, **10**(5):517-522.
36. Seoighe C: **Turning the clock back on ancient genome duplication.** *Curr Opin Genet Dev* 2003, **13**(6):636-643.
37. Patthy L: **Genome evolution and the evolution of exon-shuffling--a review.** *Gene* 1999, **238**(1):103-114.
38. Patthy L: **Modular assembly of genes and the evolution of new functions.** *Genetica* 2003, **118**(2-3):217-231.
39. Sower SSK, Reed K: **Perspective: Research activity of enteropancreatic and brain/central nervous system hormones across invertebrates and vertebrates.** *American Zoology* 2000, **40**:165-178.
40. Holland PV: **Gene duplication: past, present and future.** *Semin Cell Dev Biol* 1999, **10**(5):541-547.
41. Wang Y, Gu X: **Evolutionary patterns of gene families generated in the early stage of vertebrates.** *J Mol Evol* 2000, **51**(1):88-96.
42. Dehal P, Satou Y, Campbell RK, Chapman J, Degnan B, De Tomaso A, Davidson B, Di Gregorio A, Gelpke M, Goodstein DM, et al.: **The draft genome of Ciona intestinalis: insights into chordate and vertebrate origins.** *Science* 2002, **298**(5601):2157-2167.
43. Skrabanek L, Wolfe KH: **Eukaryote genome duplication - where's the evidence?** *Curr Opin Genet Dev* 1998, **8**(6):694-700.
44. Hughes AL: **Phylogenies of developmentally important proteins do not support the hypothesis of two rounds of genome duplication early in vertebrate history.** *J Mol Evol* 1999, **48**(5):565-576.
45. Martin WL, Pretlove S, Mercer A, Platt CC, Roberts E, Davison V, Kilby MD: **Duplication of chromosome 2 in association with ventriculomegaly - a case report.** *Prenat Diagn* 2001, **21**(13):1169-1170.
46. Lagerstrom MC, Hellstrom AR, Gloriam DE, Larsson TP, Schioth HB, Fredriksson R: **The G Protein-Coupled Receptor Subset of the Chicken Genome.** *PLoS Comput Biol* 2006, **2**(6):e54.
47. Mushegian AR, Garey JR, Martin J, Liu LX: **Large-scale taxonomic profiling of eukaryotic model organisms: a comparison of orthologous proteins encoded by the human, fly, nematode, and yeast genomes.** *Genome Res* 1998, **8**(6):590-598.
48. Coghlan A, Wolfe KH: **Fourfold faster rate of genome rearrangement in nematodes than in Drosophila.** *Genome Res* 2002, **12**(6):857-867.
49. Gonzalez J, Ranz JM, Ruiz A: **Chromosomal elements evolve at different rates in the Drosophila genome.** *Genetics* 2002, **161**(3):1137-1154.
50. Stein LD, Bao Z, Blasiar D, Blumenthal T, Brent MR, Chen N, Chinwalla A, Clarke L, Clee C, Coghlan A, et al.: **The genome sequence of Caenorhabditis briggsae: a platform for comparative genomics.** *PLoS Biol* 2003, **1**(2):E45.
51. Rogers JH: **The role of introns in evolution.** *FEBS Lett* 1990, **268**(2):339-343.
52. Rogozin IB, Wolf YI, Sorokin AV, Mirkin BG, Koonin EV: **Remarkable interkingdom conservation of intron positions and massive, lineage-specific intron loss and gain in eukaryotic evolution.** *Curr Biol* 2003, **13**(17):1512-1517.
53. Gotoh O: **Divergent structures of Caenorhabditis elegans cytochrome P450 genes suggest the frequent loss and gain of introns during the evolution of nematodes.** *Mol Biol Evol* 1998, **15**(11):1447-1459.
54. Raible F, Tessmar-Raible K, Osoegawa K, Wincker P, Jubin C, Balavoine G, Ferrier D, Benes V, de Jong P, Weissenbach J, et al.: **Vertebrate-type intron-rich genes in the marine annelid Platynereis dumerilii.** *Science* 2005, **310**(5752):1325-1326.
55. Gilbert W, Glynias M: **On the ancient nature of introns.** *Gene* 1993, **135**(1-2):137-144.
56. Foord SM: **Matching accessories.** *Sci STKE* 2003, **2003**(190):pe25.

57. Foord SM, Topp SD, Abramo M, Holbrook JD: **New methods for researching accessory proteins.** *J Mol Neurosci* 2005, **26**(2-3):265-276.
58. Parameswaran N, Spielman WS: **RAMPs: the past, present and future.** *Trends Biochem Sci* 2006, **31**(11):631-638.
59. McLatchie LM, Fraser NJ, Main MJ, Wise A, Brown J, Thompson N, Solari R, Lee MG, Foord SM: **RAMPs regulate the transport and ligand specificity of the calcitonin-receptor-like receptor.** *Nature* 1998, **393**(6683):333-339.
60. Mayo KE, Miller LJ, Bataille D, Dalle S, Goke B, Thorens B, Drucker DJ: **International Union of Pharmacology. XXXV. The glucagon receptor family.** *Pharmacol Rev* 2003, **55**(1):167-194.
61. Chow BK, Moon TW, Hoo RL, Yeung CM, Muller M, Christos PJ, Mojsos S: **Identification and characterization of a glucagon receptor from the goldfish *Carassius auratus*: implications for the evolution of the ligand specificity of glucagon receptors in vertebrates.** *Endocrinology* 2004, **145**(7):3273-3288.
62. Lefebvre PJ: **Glucagon and its family revisited.** *Diabetes Care* 1995, **18**(5):715-730.
63. Wendelaar Bonga SE, Pang PK: **Control of calcium regulating hormones in the vertebrates: parathyroid hormone, calcitonin, prolactin, and stanniocalcin.** *Int Rev Cytol* 1991, **128**:139-213.
64. Heyland A, Moroz LL: **Cross-kingdom hormonal signaling: an insight from thyroid hormone functions in marine larvae.** *J Exp Biol* 2005, **208**(Pt 23):4355-4361.
65. Hauger RL, Risbrough V, Brauns O, Dautzenberg FM: **Corticotropin releasing factor (CRF) receptor signaling in the central nervous system: new molecular targets.** *CNS Neurol Disord Drug Targets* 2006, **5**(4):453-479.
66. Cardoso JC, Power DM, Elgar G, Clark MS: **Isolation and characterisation of the corticotropin releasing factor receptor 1 (CRFRI) gene in a teleost fish, *Fugu rubripes*.** *DNA Seq* 2003, **14**(3):215-218.
67. Perrin MH, Vale WW: **Corticotropin releasing factor receptors and their ligand family.** *Ann N Y Acad Sci* 1999, **885**:312-328.
68. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Res* 1997, **25**(17):3389-3402.
69. **CMBI Blast** [<http://mrs.cmbi.ru.nl/mrs/cgi-bin/mrs-blast.cgi>]
70. **TMHMM Server v. 2.0** [<http://www.cbs.dtu.dk/services/TMHMM/>]
71. **PRINTS database** [<http://bioinf.man.ac.uk/dbbrowser/PRINTS/>]
72. **Spidey** [<http://www.ncbi.nlm.nih.gov/IEB/Research/Ostell/Spidey/>]
73. **NCBI Mapview database** [<http://www.ncbi.nlm.nih.gov/mapview/>]
74. **The Fugu Informatics Network** [<http://fugu.biology.qmul.ac.uk/>]
75. **NCBI Blast server** [<http://www.ncbi.nlm.nih.gov/blast>]
76. Thompson JD, Higgins DG, Gibson TJ: **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.** *Nucleic Acids Res* 1994, **22**(22):4673-4680.
77. **GeneDoc** [<http://www.psc.edu/biomed/genedoc>]
78. Kumar S, Tamura K, Nei M: **Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment.** *Brief Bioinform* 2004, **5**(2):150-163.
79. Giribet G: **Relationships among metazoan phyla as inferred from 18S rRNA sequence data: a methodological approach.** *EXS* 2002:85-101.
80. Gerhart J, Lowe C, Kirschner M: **Hemichordates and the origin of chordates.** *Curr Opin Genet Dev* 2005, **15**(4):461-467.
81. Delsuc F, Brinkmann H, Chourrout D, Philippe H: **Tunicates and not cephalochordates are the closest living relatives of vertebrates.** *Nature* 2006, **439**(7079):965-968.
82. Blair Hedges S, Kumar S: **Genomic clocks and evolutionary timescales.** *Trends Genet* 2003, **19**(4):200-206.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

